

Ethnic variations in the occurrence of colonic neoplasms

Amnon Sonnenberg^{1,2}, Kevin O Turner¹ and Robert M Genta^{1,3}

Abstract

Background and aims: With the exception of African Americans and Hispanics, few studies have dealt with the influence of other types of ethnicity on the prevalence of colon polyps and colorectal cancer. The present study was undertaken to compare the ethnic and socioeconomic distributions of colonic neoplasms among different ethnic groups in the United States.

Methods: A total of 813,057 patients, who underwent colonoscopy during 2008–2014, were recruited from an electronic database of histopathology reports (Miraca Life Sciences) for a cross-sectional study. Using multivariate logistic regression analyses, the presence of hyperplastic polyps, serrated adenomas, tubular adenomas, or adenocarcinomas each served as separate outcome variables. Patient ethnicity was determined using a name-based computer algorithm. Demographic (age, sex, ethnicity) and a variety of socioeconomic risk factors (associated with patients' ZIP code) served as predictor variables.

Results: About 50% of the study population harbored adenomatous polyps, 25% hyperplastic polyps, 8% serrated adenomas, and 1.4% adenocarcinomas. Tubular adenomas and adenocarcinomas showed similar ethnic distributions, being slightly more common among Hispanics and East Asians. All four types of colonic neoplasm were relatively rare among patients of Asian-Indian descent and relatively common among patients of Japanese descent. Except for Japanese patients, serrated adenomas tended to be less prevalent among East Asians. In general, markers of high socioeconomic status showed a tendency to be negatively associated with the presence of tubular adenoma and adenocarcinoma, but positively with the presence of serrated adenoma.

Conclusion: Ethnicity and socioeconomic factors affect different histology types of polyps differently. Genetic as well as environmental factors interact in the development of colorectal cancer and its precursor lesions.

Keywords

Colon polyps, colorectal cancer, demographic characteristics, environmental epidemiology, socioeconomic risk factors

Received: 28 June 2016; accepted: 9 August 2016

Introduction

The ethnic variation in the occurrence of benign and malignant colonic neoplasms within different ethnic subgroups in the United States (US) population has continued to capture the interest of gastroenterologists and epidemiologists alike. Previous studies have suggested that the occurrence of colon polyps and colorectal cancer are more common among African Americans than Whites.^{1–5} While some studies have shown a decreased incidence and mortality of colorectal cancer among Hispanics,⁶ other studies have found decreased,⁴ similar,^{3,7} or even increased¹ prevalence rates of polyps among Hispanics when compared to Whites. Data from the Surveillance, Epidemiology, and End Results (SEER) program and the Vital Statistics of the US suggested a decreased incidence

and mortality, respectively, of colorectal cancer among US residents of Asian descent.⁵ Previous studies have been partly limited by small study populations and lack of stratification by histopathology.

The Miraca Life Sciences Database is an electronic repository of histopathologic patient records. Biopsy specimens are submitted to Miraca Life Sciences by approximately 1500 gastroenterologists from private

¹Miraca Life Sciences, Irving, TX, USA

²Oregon Health & Science University, Portland, OR, USA

³Baylor College of Medicine, Houston, TX, USA

Corresponding author:

Amnon Sonnenberg, Portland VA Medical Center, Department of Gastroenterology, 3710 SW US Veterans Hospital Road, P3-G1, Portland, OR 97239, USA.

Email: sonnenbe@ohsu.edu

practices distributed throughout the US. Each annual file contains the records of more than 200,000 patients who have undergone upper or lower gastrointestinal endoscopy with tissue biopsies. In the recent past, this database has been used for a variety of pathoepidemiologic studies.⁸⁻¹⁴ The aim of the present analysis was to use this unique database to study the ethnic variation in the occurrence of various types of colonic neoplasms.

Material and methods

Data source

The Miraca Life Sciences database contains demographic information and a detailed list of all results of surgical pathology reports, which are coded in a pre-defined, standardized, and searchable fashion. The database was searched for the records of all colonoscopies performed between January 2008 and December 2014. If a patient had multiple colonoscopies, only data from the chronologically first procedure were included. Colonoscopies were selected irrespective of their primary indication, such as screening, surveillance of known disease, or workup of new symptoms. In patients with colonic neoplasms, the neoplasms were categorized as hyperplastic polyps, sessile serrated adenomas, tubular adenomas, or adenocarcinomas. An individual patient with simultaneous occurrence of multiple neoplasms could thus contribute to multiple categories. Social history, medication list, and results of laboratory tests are infrequently provided with the pathology specimen and were, therefore, not used for the purpose of this study.

Unique patients with colonic tissue specimens were extracted from the database, and their demographic and histopathologic data were recorded. Socioeconomic information was available from census data associated with the patients' postal address and Zoning Improvement

Plan (ZIP) code. These data included population size per ZIP code, average number of people per household, average house value, average annual income, and percentage of residents with college education. Small population sizes may reflect on more exclusive residential areas. Based on complex computer algorithms, using first and last names, patients were also grouped by ethnicity as follows: Hispanic, Indian (Indian subcontinent), East Asian (Chinese, Korean, Japanese, and Vietnamese), Portuguese, Jewish, Arab, and other. The latter group included US residents (mostly Caucasians and African Americans) and patients not identified with any of the above groups. The algorithm has been explained in greater detail in a recent publication.¹⁴

All data were derived from pre-existing records. No direct contact with either patients or providers was made and no individual patient information was revealed. All patient records were de-identified before being analyzed. For these reasons, the study protocol was exempted from the need for informed consent from its participants.

Statistical analyses

Patients were stratified by ethnicity and presence or absence of various types of colonic neoplasm. The prevalence of each type of neoplasm was expressed as a percentage of all patients belonging to the same ethnic category. In four different multivariate logistic regression analyses, the presence of hyperplastic polyps, serrated adenomas, tubular adenomas, or adenocarcinomas each served as separate outcome variables. Age, sex, ethnicity, population size, people per household, house value, annual income, and college education served as predictor variables. The influences of the predictors on each outcome were expressed as odds ratios (ORs) with their 95% confidence intervals (CIs). For continuous predictor variables, such as age, population size, and

Table 1. Patient distribution by ethnicity and type of colonic neoplasm

Ethnicity	Hyperplastic polyp		Serrated adenoma		Tubular adenoma		Adeno-carcinoma		Grand total
Total	207,540	(26%)	68,189	(8%)	404,343	(50%)	11,167	(1.4%)	813,057
Hispanic	8887	(21%)	2372	(5%)	22,997	(53%)	679	(1.6%)	43,243
Indian	359	(17%)	80	(4%)	912	(42%)	23	(1.1%)	2150
Japanese	823	(32%)	234	(9%)	1397	(54%)	38	(1.5%)	2582
Korean	975	(24%)	223	(6%)	2189	(55%)	69	(1.7%)	3984
Vietnamese	498	(22%)	101	(4%)	1347	(59%)	40	(1.7%)	2302
Chinese	1400	(24%)	343	(6%)	3092	(53%)	97	(1.7%)	5845
Portuguese	1165	(28%)	308	(7%)	2228	(54%)	52	(1.3%)	4130
Jewish	7578	(24%)	2680	(9%)	14,846	(48%)	368	(1.2%)	30,953
Arab	251	(18%)	54	(4%)	671	(48%)	15	(1.1%)	1387

house values, the OR and CI were expressed per change over the entire range.

Results

The database contained a total of 813,057 patients who underwent colonoscopy. Table 1 shows a stratification of the patient population by histologic type and patient ethnicity. The percentage values represent prevalence for each ethnic group. (The values in each row do not all add up to the grand total because of non-polyp biopsy findings in the entire patient population. Individual patients could also present with more than one polyp type.) Only a minority of patients belonged to any of the nine predefined ethnic groups. Hispanics and Jews constituted the two largest groups with 43,000 and 31,000 patients, respectively. The size of the five separate Asian groups ranged between 2300 and 5800 patients. Patients of Arab descent constituted the smallest ethnic subgroup.

About 50% of the population harbored adenomatous polyps, 25% hyperplastic polyps, 8% serrated adenomas, and 1.4% adenocarcinomas. Figure 1 contains the prevalence of the four histologic types by ethnic subgroup. Several general patterns were discernible. All four types of colonic neoplasm were relatively rare among Americans of Indian and Arab descent and relatively common among Americans of Japanese descent. Tubular adenomas and adenocarcinomas showed a similar ethnic distribution. They both tended to be slightly more common among Hispanics and East Asians. Except for Japanese Americans, serrated adenomas tended to be less prevalent among all East Asians.

With the notable exception of serrated adenomas, all types of colon neoplasms were more common in men than women (Figure 2). The prevalence of all types of colon neoplasm showed an age-dependent rise (Figure 3). This rise was most pronounced in tubular adenoma and colorectal cancer. The age-specific prevalence of serrated adenomas and hyperplastic polyps increased between the age groups 0–9 and 50–59 years and subsequently showed a smooth decline.

Tables 2–5 contain the results of the multivariate logistic regressions. Besides type of ethnicity, age, and sex, population size per ZIP code, average number of people per household, average house value, annual income, and percentage of college education per ZIP code served as additional predictor variables. ORs are

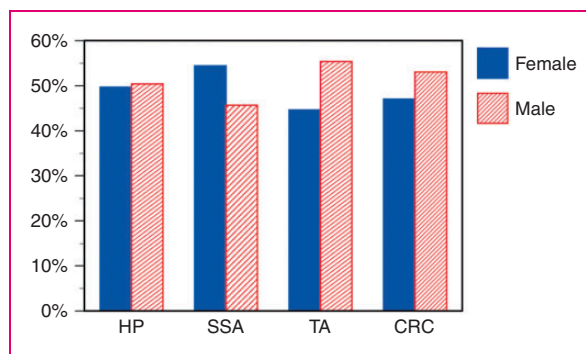


Figure 2. Gender distribution of hyperplastic polyps (HP), sessile serrated adenomas (SSA), tubular adenomas (TA), and colorectal cancer (CRC).

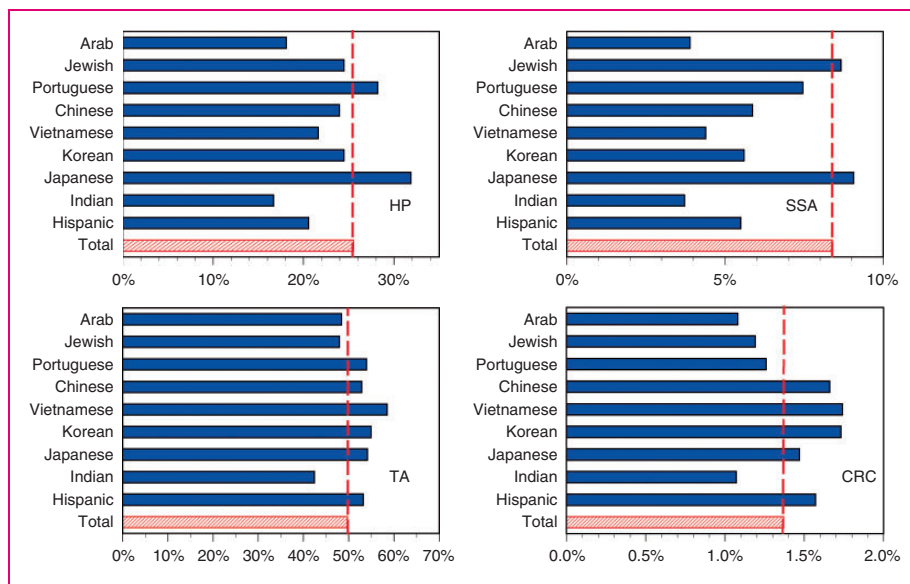


Figure 1. Prevalence of hyperplastic polyps (HP), sessile serrated adenomas (SSA), tubular adenomas (TA), or colorectal cancer (CRC) in different ethnic groups.

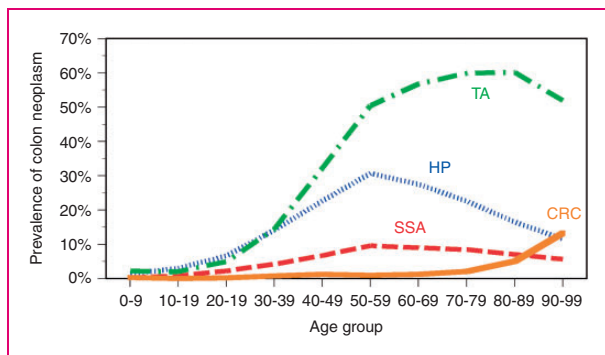


Figure 3. Age distribution of hyperplastic polyps (HP), sessile serrated adenomas (SSA), tubular adenomas (TA), and colorectal cancer (CRC) in different ethnic groups.

Table 2. Logistic fit for hyperplastic polyps

Variable	OR ^a	95%	Chi-square	Probability
Age	1.53	(1.47-1.59)	432.72	<0.0001
Males	1.12	(1.11-1.13)	517.96	<0.0001
Population size	0.92	(0.89-0.95)	20.12	<0.0001
People per household	1.06	(0.99-1.14)	3.20	0.0737
House value	1.03	(0.98-1.08)	1.22	0.2688
Annual income	1.29	(1.17-1.41)	28.02	<0.0001
College education	0.73	(0.69-0.78)	107.75	<0.0001
Hispanic	0.74	(0.72-0.76)	583.25	<0.0001
Indian	0.59	(0.53-0.66)	83.09	<0.0001
Japanese	1.32	(1.21-1.43)	41.11	<0.0001
Korean	0.95	(0.88-1.02)	1.80	0.1796
Vietnamese	0.80	(0.72-0.88)	19.51	<0.0001
Chinese	0.93	(0.87-0.99)	5.12	0.0237
Portuguese	1.15	(1.08-1.24)	17.06	<0.0001
Jewish	0.93	(0.91-0.96)	26.12	<0.0001
Arab	0.64	(0.56-0.74)	40.19	<0.0001

^aOdds ratio (OR) and 95% confidence interval (CI) per change in predictor variable over its entire range.

shown for the entire range of each predictor variable. In addition to the ORs, their associated chi-square values may also provide estimates for the strength of the individual relationships. For all polyps and adenocarcinoma alike, age represented the most influential predictor variable. Its effect was most pronounced in tubular adenoma and adenocarcinoma, much less so in serrated adenoma, and least in hyperplastic polyps. With the notable exception of serrated adenomas, all colonic neoplasms were more common in men than women.

The relationships between the ZIP-associated socio-demographic parameters and the prevalence of colonic neoplasia revealed multiple significant associations.

Table 3. Logistic fit for serrated adenomas

Variable	OR ^a	95%	Chi-square	Probability
Age	2.46	(2.31-2.62)	763.23	<0.0001
Males	0.88	(0.87-0.90)	229.95	<0.0001
Population size	0.80	(0.76-0.85)	55.17	<0.0001
People per household	0.93	(0.83-1.03)	1.81	0.1786
House value	0.67	(0.62-0.72)	109.76	<0.0001
Annual income	1.55	(1.34-1.80)	34.83	<0.0001
College education	1.90	(1.74-2.08)	196.40	<0.0001
Hispanic	0.68	(0.65-0.71)	310.71	<0.0001
Indian	0.42	(0.34-0.53)	57.67	<0.0001
Japanese	1.13	(0.99-1.30)	3.33	0.0678
Korean	0.70	(0.61-0.80)	25.70	<0.0001
Vietnamese	0.53	(0.43-0.64)	39.55	<0.0001
Chinese	0.74	(0.67-0.83)	26.73	<0.0001
Portuguese	0.98	(0.87-1.10)	0.12	0.7268
Jewish	0.97	(0.94-1.01)	1.57	0.2100
Arab	0.47	(0.36-0.62)	28.67	<0.0001

^aOdds ratio (OR) and 95% confidence interval (CI) per change in predictor variable over its entire range.

Table 4. Logistic fit tubular adenomas

Variable	OR ^a	95%	Chi-square	Probability
Age	69.98	(67.29-72.77)	45327.00	<0.0001
Males	1.79	(1.78-1.81)	15769.00	<0.0001
Population size	0.88	(0.86-0.91)	53.37	<0.0001
People per household	2.70	(2.54-2.87)	1011.80	<0.0001
House value	0.83	(0.79-0.86)	74.86	<0.0001
Annual income	0.50	(0.46-0.54)	258.70	<0.0001
College education	1.46	(1.39-1.54)	199.53	<0.0001
Hispanic	1.12	(1.09-1.14)	106.77	<0.0001
Indian	0.81	(0.74-0.89)	20.19	<0.0001
Japanese	1.13	(1.04-1.23)	9.01	0.0027
Korean	1.14	(1.07-1.22)	14.69	0.0001
Vietnamese	1.39	(1.28-1.52)	56.39	<0.0001
Chinese	1.14	(1.08-1.21)	22.49	<0.0001
Portuguese	1.19	(1.12-1.27)	29.58	<0.0001
Jewish	0.92	(0.90-0.94)	48.73	<0.0001
Arab	0.96	(0.86-1.07)	0.47	0.4943

^aOdds ratio (OR) and 95% confidence interval (CI) per change in predictor variable over its entire range.

The average number of people per household may be interpreted as an indirect marker for poverty or low socioeconomic status, whereas average income is indicative of affluence or high socioeconomic status. A high average number of people per household was positively associated both with tubular adenoma and

Table 5. Logistic fit for adenocarcinoma

Variable	OR ^a	95%	Chi-square	Probability
Age	150.76	(127.05-178.99)	3292.70	<0.0001
Males	1.20	(1.15-1.24)	88.06	<0.0001
Population size	1.14	(1.00-1.31)	3.86	0.0495
People per household	1.74	(1.37-2.20)	21.24	<0.0001
House value	1.19	(0.99-1.42)	3.59	0.0580
Annual income	0.57	(0.40-0.82)	9.41	0.0022
College education	0.70	(0.56-0.88)	9.64	0.0019
Hispanic	1.08	(0.99-1.17)	3.27	0.0705
Indian	0.93	(0.61-1.40)	0.13	0.7222
Japanese	0.96	(0.70-1.33)	0.05	0.8240
Korean	1.09	(0.85-1.40)	0.45	0.5046
Vietnamese	1.25	(0.91-1.72)	1.95	0.1630
Chinese	1.20	(0.97-1.48)	2.84	0.0917
Portuguese	0.88	(0.67-1.16)	0.87	0.3501
Jewish	0.87	(0.78-0.96)	7.20	0.0073
Arab	0.88	(0.53-1.46)	0.25	0.6160

^aOdds ratio (OR) and 95% confidence interval (CI) per change in predictor variable over its entire range.

adenocarcinoma, whereas high income was inversely associated both with tubular adenoma and adenocarcinoma. College education is generally also indicative of affluence or high socioeconomic status, but it was inversely associated only with adenocarcinoma and not tubular adenoma. Serrated adenomas seemed to behave oppositely to tubular adenoma and adenocarcinoma in that markers of high socioeconomic status, such as annual income and college education, were both positively associated with its prevalence.

The multivariate analyses largely confirmed the ethnic association described above, based on the prevalence rates. Being of Hispanic or East Asian origin (Japanese, Korean, Vietnamese, or Chinese), was associated with an increased OR for harboring tubular adenomas. Although similar relationships also applied to colonic adenocarcinoma, because of low case numbers, they failed to reach statistical significance for colorectal cancer in individual ethnic groups. With all East Asians combined, however, the OR reached statistical significance at OR = 1.15 (95% CI: 1.02–1.33). Jewish ancestry was inversely associated both with tubular adenoma and adenocarcinoma. As noted above, with the exception of Japanese Americans, serrated adenomas were inversely associated with East Asian descent. Being Hispanic was also associated with a low OR for serrated adenomas.

The ZIP-associated socio-demographic factors exerted a much weaker influence on hyperplastic polyps than on the other types of colonic neoplasia. As in other types of colonic neoplasms, the presence of hyperplastic polyps

was negatively associated with Indian and positively associated with Japanese descent. Otherwise, hyperplastic polyps tended to be inversely associated with East Asian (except Japanese) and Hispanic ethnicity.

Discussion

The present study was undertaken to compare the ethnic and socioeconomic distributions of colonic neoplasms among different groups in the US. Our study showed that tubular adenomas and adenocarcinomas were characterized by similar ethnic distributions, being slightly more common among Hispanics and East Asians. Except for Japanese Americans, serrated adenomas tended to be less prevalent among all East Asians. In general, markers of high socioeconomic status showed a tendency to be negatively associated with the presence of tubular adenoma and adenocarcinoma, but positively with the presence of serrated adenoma.

There are only a few studies that have dealt with the influence of ethnicity on the prevalence of colon polyps and colorectal cancer. The majority of studies have focused on the occurrence of colonic neoplasms in African Americans and, to a lesser extent, Hispanics. Most studies have reported a higher prevalence of polyps in African Americans when compared to Whites,¹⁻⁴ as well as higher incidence and mortality rates of colonic adenocarcinoma.⁵ Higher incidence and mortality rates among African Americans seem to be also influenced by issues of access to and utilization of health care by the African American population.¹⁵⁻¹⁷ Besides ethnicity, socioeconomic factors have been shown to affect incidence, mortality, and survival of colorectal cancer patients.¹⁸ Fewer studies have included a sufficient number of Hispanics for statistical analysis. In contradistinction with African Americans, the issue of colon neoplasm prevalence in Hispanics relatively to Whites is less well settled. Incidence and mortality of colorectal cancer appear to be slightly less than in Whites,⁶ whereas the prevalence of polyps was found to be lower, as well as equal or even higher in Hispanics than in Whites.^{1,3,4,7,19} Lastly, based on few available sources, the prevalence of polyps among Asian Americans appears to be equal to those of Whites, whereas incidence and mortality of colorectal cancer appear to be slightly less than those of White Americans.^{3,5}

In general, the differences between Hispanic or Asians and Whites are smaller than the difference between African Americans and Whites. Part of the problem in establishing the relative frequency of colonic neoplasm in ethnicities other than White and African American relates to the relatively small sample sizes available for patients of different ethnic background. In addition, the terms “Asian” and “Hispanic” include a variety

of ethnically diverse people with very different cultural and genetic backgrounds. In many existing studies, a population consisting 90% of Chinese and 10% of Indian individuals would be listed as Asian, as would one consisting of 90% Indians and 10% Chinese. In these two populations many gastrointestinal conditions (e.g. preneoplastic gastric lesions²⁰) would be extremely different and the resulting data would be erroneously interpreted as discrepant. Our method of separating patients in highly specific ethnic groups largely avoids this pitfall, except in the case of Hispanics, who share a similar pool of names in Spain and all Spanish-speaking countries in South and Central America.

Lastly, the ethnic distribution may also vary in different types of colonic neoplasm. In the present study, tubular adenomas and adenocarcinomas were characterized by resembling patterns, probably reflecting the fact that the majority of adenocarcinomas originate from tubular adenomas. Both were slightly more common in Hispanics and East Asians, although the positive associations between colonic adenocarcinoma and Asians failed to reach statistical significance for each individual ethnic group analyzed separately. As a composite group, however, East Asian ethnicity was associated with a slightly but significantly higher prevalence of adenocarcinoma. There were also notable differences with the broad category of Asian Americans. Japanese patients were characterized by consistently high prevalence rates for all colon neoplasms, whereas Indian patients were characterized by consistently low prevalence rates. To complicate matters even further, serrated adenomas behaved differently from tubular adenomas. The prevalence of tubular adenomas was relatively high among East Asians, whereas the prevalence of serrated adenomas was relatively low. In general, serrated adenomas were influenced by other risk factors than tubular adenomas. Markers of high socioeconomic status tended to be negatively associated with tubular adenoma and adenocarcinoma, but positively with serrated adenoma.

Our study has several potential limitations. Because the primary data source was a pathology database, we had little if any access to any detailed information about a multitude of other risk factors, such as comorbidities, social or dietary habits, which have been found to influence the occurrence of colonic neoplasm. For instance, increased consumption of alcohol, animal fat and meat, and decreased consumption of milk and calcium have all been associated with an elevated risk for colonic neoplasm.^{21–23} Diabetes, obesity, gallstone disease, and smoking also increase the risk, whereas physical activity and the intake of nonsteroidal anti-inflammatory drugs decrease the risk.^{23–26}

Data on socioeconomic status were based on patients' place of residence, as evidenced by the ZIP

code, rather than personal information. Many socioeconomic parameters are no longer entered into the medical record because they relate to sensitive or confidential information. To circumvent such potential limitations, it has become a common practice in epidemiologic research to use socioeconomic data available from a patient's ZIP code as an alternative marker of socioeconomic status.^{27–29} However, more precise and personalized information would have probably accentuated the socioeconomic patterns observed by the present analysis.

In spite of the large patient population, in some instances, the number of cases per individual ethnic group still remained small. Because of the nature of this database, the individual ethnic groups were compared by their relative prevalence of colonic neoplasms. The true prevalence could not be calculated since patients without colonoscopy or patients without tissue samples would not be included in this histopathology database. These limitations are likely to have affected various ethnic groups alike. It cannot be ruled out, however, that the observed ethnic variations were also influenced by underlying variations in access to health care in general or colonoscopy in particular.

Computer algorithms to identify ethnicity by name have become widely used tools in health care research, anthropology, sociology, and other population studies.^{30–38} The technique has been shown to be characterized by high accuracy in correctly assessing ethnicity, with positive predictive values exceeding 95%. The presence of unidentified patients of Hispanic, Asian, or any other ethnicity within the comparison group of other Americans would have biased our statistical analyses toward the null hypothesis. Because our algorithms cannot be used to determine African Americans' ancestry, we did not further stratify our comparison population.

These potential limitations of the database analysis need to be contrasted with its obvious advantages. One strength of our study relates to its large population of patients who were recruited from endoscopy practices distributed throughout the entire US. The present study included more ethnic subgroups than any previous study. Because the analysis relied on endoscopic and histopathologic findings by board-certified pathologists subspecialized in gastrointestinal pathology, the ascertainment of diagnoses can be assumed to be highly reliable. Different from previous analysis, we were able to stratify colon polyps into different subtypes. Moreover, the size of the database provided the opportunity to include an appreciable number of patients with colorectal cancer.

In conclusion, the present study confirms that prevalence of colonic neoplasms is influenced by patient ethnicity and socioeconomic risk factors. Ethnicity and socioeconomic factors affect different histology types

of polyps differently. Genetic as well as environmental factors interact in the development of colorectal cancer and its precursor lesions.

Acknowledgments

Author contributions include the following: Study conception and design: RM Genta, A Sonnenberg and KO Turner; data analysis: A Sonnenberg and RM Genta; writing of manuscript: A Sonnenberg and RM Genta.

Conflict of interest

A Sonnenberg has no conflict of interest to declare. KO Turner and RM Genta are employed by Miraca Life Sciences, Irving, TX.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

- Lebwohl B, Capiak K, Neugut AI, et al. Risk of colorectal adenomas and advanced neoplasia in Hispanic, black and white patients undergoing screening colonoscopy. *Aliment Pharmacol Ther* 2012; 35: 1467–1473.
- Schroy PC 3rd, Coe A, Chen CA, et al. Prevalence of advanced colorectal neoplasia in white and black patients undergoing screening colonoscopy in a safety-net hospital. *Ann Intern Med* 2013; 159: 13–20.
- Corley DA, Jensen CD, Marks AR, et al. Variation of adenoma prevalence by age, sex, race, and colon location in a large population: Implications for screening and quality programs. *Clin Gastroenterol Hepatol* 2013; 11: 172–180.
- Lieberman DA, Williams JL, Holub JL, et al. Race, ethnicity, and sex affect risk for polyps >9 mm in average-risk individuals. *Gastroenterology* 2014; 147: 351–358. quiz e14–e15.
- Siegel R, Naishadham D and Jemal A. Cancer statistics, 2012. *CA Cancer J Clin* 2012; 62: 10–29.
- Siegel R, Naishadham D and Jemal A. Cancer statistics for Hispanics/Latinos, 2012. *CA Cancer J Clin* 2012; 62: 283–298.
- Thoma MN, Jimenez Cantisano BG, Hernandez AV, et al. Comparison of adenoma detection rate in Hispanics and whites undergoing first screening colonoscopy: A retrospective chart review. *Gastrointest Endosc* 2013; 77: 430–435.
- Sonnenberg A, Lash RH and Genta RM. A national study of *Helicobacter pylori* infection in gastric biopsy specimens. *Gastroenterology* 2010; 139: 1894–1901.
- Dellon ES, Peery AF, Shaheen NJ, et al. Inverse association of esophageal eosinophilia with *Helicobacter pylori* based on analysis of a US pathology database. *Gastroenterology* 2011; 141: 1586–1592.
- Gupta S, Balasubramanian BA, Fu T, et al. Polyps with advanced neoplasia are smaller in the right than in the left colon: Implications for colorectal cancer screening. *Clin Gastroenterol Hepatol* 2012; 10: 1395–1401.
- Sonnenberg A and Genta RM. *Helicobacter pylori* is a risk factor for colonic neoplasms. *Am J Gastroenterol* 2013; 108: 208–215.
- Lebwohl B, Blaser MJ, Ludvigsson JF, et al. Decreased risk of celiac disease in patients with *Helicobacter pylori* colonization. *Am J Epidemiol* 2013; 178: 1721–1730.
- Jensen ET, Hoffman K, Shaheen NJ, et al. Esophageal eosinophilia is increased in rural areas with low population density: Results from a national pathology database. *Am J Gastroenterol* 2014; 109: 668–675.
- Turner KO, Genta RM and Sonnenberg A. Ethnic distribution of microscopic colitis in the United States. *Inflamm Bowel Dis* 2015; 21: 2634–2639.
- Laiyemo AO, Doubeni C, Pinsky PF, et al. Race and colorectal cancer disparities: Health-care utilization vs different cancer susceptibilities. *J Natl Cancer Inst* 2010; 102: 538–546.
- Laiyemo AO, Doubeni C, Pinsky PF, et al. Occurrence of distal colorectal neoplasia among whites and blacks following negative flexible sigmoidoscopy: An analysis of PLCO Trial. *J Gen Intern Med* 2015; 30: 1447–1453.
- Bromley EG, May FP, Federer L, et al. Explaining persistent under-use of colonoscopic cancer screening in African Americans: A systematic review. *Prev Med* 2015; 71: 40–48.
- Manser CN and Bauerfeind P. Impact of socioeconomic status on incidence, mortality, and survival of colorectal cancer patients: A systematic review. *Gastrointest Endosc* 2014; 80: 42–60.
- Lee KK, Jandorf L, Thélèmanque L, et al. Colorectal neoplasia detection among black and Latino individuals undergoing screening colonoscopy: A prospective cohort study. *Gastrointest Endosc* 2014; 79: 466–472.
- Choi CE, Sonnenberg A, Turner K, et al. High prevalence of gastric preneoplastic lesions in East Asians and Hispanics in the USA. *Dig Dis Sci* 2015; 60: 2070–2076.
- Ferrari P, Jenab M, Norat T, et al. Lifetime and baseline alcohol intake and risk of colon and rectal cancers in the European prospective investigation into cancer and nutrition (EPIC). *Int J Cancer* 2007; 121: 2065–2072.
- Cho E, Smith-Warner SA, Spiegelman D, et al. Dairy foods, calcium, and colorectal cancer: A pooled analysis of 10 cohort studies. *J Natl Cancer Inst* 2004; 96: 1015–1022.
- Huxley RR, Ansary-Moghaddam A, Clifton P, et al. The impact of dietary and lifestyle risk factors on risk of colorectal cancer: A quantitative overview of the epidemiological evidence. *Int J Cancer* 2009; 125: 171–180.
- Rahman MI, Gibson-Shreve LD, Yuan Z, et al. Selections from current literature: Cholelithiasis, cholecystectomy and the risk of colorectal cancer. *Fam Pract* 1996; 13: 483–487.
- Samad AK, Taylor RS, Marshall T, et al. A meta-analysis of the association of physical activity with reduced risk of colorectal cancer. *Colorectal Dis* 2005; 7: 204–213.
- Rothwell PM, Wilson M, Elwin CE, et al. Long-term effect of aspirin on colorectal cancer incidence and

- mortality: 20-year follow-up of five randomised trials. *Lancet* 2010; 376: 1741–1750.
27. Krieger N. Overcoming the absence of socioeconomic data in medical records: Validation and application of a census-based methodology. *Am J Public Health* 1992; 82: 703–710.
 28. Krieger N, Williams D and Moss N. Measuring social class in US public health research: Concepts, methodologies and guidelines. *Annu Rev Public Health* 1997; 18: 341–347.
 29. Krieger N, Chen JT, Waterman PD, et al. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: Does the choice of area-based measure and geographic level matter? *Am J Epidemiol* 2002; 156: 471–482.
 30. Fiscella K and Fremont AM. Use of geocoding and surname analysis to estimate race and ethnicity. *Health Serv Res* 2006; 41(4 Pt 1): 1482–1500.
 31. Mateos P. A review of name-based ethnicity classification methods and their potential in population studies. *Popul Space Place* 2007; 13: 243–263.
 32. Coldman AJ, Braun T and Gallagher RP. The classification of ethnic status using name information. *J Epidemiol Community Health* 1988; 42: 390–395.
 33. Lakha F, Gorman DR and Mateos P. Name analysis to classify populations by ethnicity in public health: Validation of Onomap in Scotland. *Public Health* 2011; 125: 688–696.
 34. Lauderdale DS and Kestenbaum B. Asian American ethnic identification by surname. *Popul Res Policy Rev* 2000; 19: 283–300.
 35. Macfarlane GJ, Lunt M, Palmer B, et al. Determining aspects of ethnicity amongst persons of South Asian origin: The use of a surname-classification programme (Nam Pehchan). *Public Health* 2007; 121: 231–236.
 36. Nitsch D, Kadalayil L, Steenkamp R, et al. Validation and utility of a computerized South Asian names and group recognition algorithm in ascertaining South Asian ethnicity in the national renal registry. *Q J Med* 2009; 102: 865–872.
 37. Quan H, Wang F, Schopflocher D, et al. Development and validation of a surname list to define Chinese ethnicity. *Med Care* 2006; 44: 328–333.
 38. Ryan R, Vernon S, Lawrence G, et al. Use of name recognition software, census data and multiple imputation to predict missing data on ethnicity: Application to cancer registry records. *BMC Med Inform Decis Mak* 2012; 12: 3.